DOCUMENT RESUME

ED 161 891 TM 007 199

AUTHOR Skakun, Ernest N.; Hakstian, A. Ralph
TITLE An Examination of Sampling Characteristics of Some Analytic Factor Transformation Techniques.
INSTITUTION Alberta Univ., Edmonton. Div. of Educational Research.
REPORT NO RIR-74-1
PUB DATE Mar 74
NOTE 12p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Comparative Statistics; *Factor Analysis; *Matrices; Models; *Oblique Rotation; *Orthogonal Rotation; *Sampling; Simulation; *Statistical Analysis
IDENTIFIERS Least Squares Analysis

ABSTRACT

Two population raw data matrices were constructed by computer simulation techniques. Each consisted of 10,000 subjects and 12 variables, and each was constructed according to an underlying factorial model consisting of four major common factors, eight minor common factors, and 12 unique factors. The computer simulation techniques were employed to generate one population of factorially simple, and one of complex data, with first the oblique and orthogonal factor models and then the solutions based on several special cases of the general orthomax criterion compared on these characteristics. Results indicated that a significantly better fit to the oblique population factor pattern matrix was realized by the oblique, as opposed to orthogonal, least squares transformations. When the magnitude of the differences between population values was compared with the magnitude of the mean differences for the sample values, the minute population differences had no effect on mean differences for the factorially simple data or the complex data. The authors outline the practical implications of these findings and offer recommendations for future research. (Author/JAC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Division of Educational Research Services

University of Alberta

ED161891

RIR-74-1

AN EXAMINATION OF SAMPLING CHARACTERISTICS OF SOME ANALYTIC FACTOR TRANSFORMATION TECHNIQUES

ERNEST N. SKAKUN and A. RALPH HAKSTIAN

University of Alberta

University of British Columbia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

E.W. Romaniuk

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

March, 1974

The contents of this report are not to be cited without permission of the authors.

- TM007 199 -

ABSTRACT

Given a population simple structure configuration for a set of data, factor transformation methods can be compared in terms of their ability--with samples drawn from this population--to closely approximate the parameter pattern, and to do so with little sampling variability. A study is reported, in which computer simulation techniques were employed to generate a population of factorially simple, and one of complex, data, with first the oblique and orthogonal factor models and then solutions based on several special cases of the general orthomax criterion compared on these characteristics. Results are discussed, and certain practical implications noted.

AN EXAMINATION OF SAMPLING CHARACTERISTICS OF SOME ANALYTIC FACTOR TRANSFORMATION TECHNIQUES

ERNEST N. SKAKUN
University of Alberta
and
A. RALPH HAKSTIAN
University of British Columbia

The large number of analytic transformation methods in factor analysis have, over the years, frequently been compared. Beginning with Carroll's (1953) comparison of the then-new orthogonal transformation method, quartimax, with graphic and analytic oblique solutions, and continuing to the recent comparative studies by Hakstian (1971) and Hakstian and Boyd (1972), the vast majority of studies examing factor transformation methods have presented comparisons of solutions in terms of relative closeness to an ideal (and usually known) simple structure resolution for the data. Thus, the determination of the optimal procedure centered entirely around closeness to an optimal linear transformation of a fixed (typically sample) matrix. Problems of statistical inference were ignored, and the comparative problem was exclusively algebraic, rather than statistical.

Another question can be asked, however, when comparing factor transformation methods. This is the extent to which the methods compared exhibit invariance over samples, which we define, somewhat loosely, as the tendency of a method--when applied to a sample unrotated pattern matrix--to approximate a population parameter pattern matrix. Furthermore, and in the same connection, we can inquire as to the sampling variability of various transformation methods. In general, the closer the fit between sample and population transformations, and the less the variability, the better.

In the present study, two kinds of comparative analyses were performed, both involving a large number of samples. In the first, solutions that were in a sense

"ideal" were obtained under both the oblique and orthogonal factor models. These solutions--transformed under the constraints implied in the respective models--were least squares fits to a known "true" primary-factor pattern in the population, and were obtained to assess the relative merits of the oblique, as opposed to orthogonal, model, with regard to sampling stability. In the second set of comparative analyses, solutions obtained by various "blind" orthogonal techniques were compared in terms of (1) overall (mean) closeness to the population pattern, and (2) sampling variability of the obtained solutions.

METHOD

Two population raw data matrices were constructed by computer simulation techniques. Each consisted of 10,000 "subjects" and 12 "variables," and each was constructed according to an underlying factorial model consisting of (1) four major common factors, (2) eight minor common factors, and (3) 12 unique factors. A simple structure configuration was embodied in the four major factors, and, thus, a 12 x 4 simple structure primary-factor pattern was the locus of interest in both populations. In one population--the "simple" data--the 12 variables were strictly factorially simple in terms of the four major factors. In the other--the "complex" data--six of the 12 variables were made factorially complex. In both cases, the 12 common factors were built in with a certain degree of mutual intercorrelation in the population, with a mean population interfactor r of .40 among the four major oblique simple structure factors.

At this point, 100 samples, each consisting of 100 "subjects" were drawn with replacement from each population. For each sample, the first four principal components were extracted, yielding 100 F matrices, each 12 x 4, for both the factorially simple and complex populations. Thus, at this point, a discrepancy was established between the underlying (common-factor) model in the populations and the (component) model implied by the factoring procedure. The latter model was used because of (1) its widespread use in practice, and (2) its greater economy, an important

consideration in a study of this magnitude. A 12 x 4 factor pattern for each of the two entire populations was obtained and transformed to an oblique simple structure pattern, $\underline{P}^*$. These two component pattern matrices, $\underline{P}^*$--one for the simple and one for the complex data--were transformed to as close as possible a fit to the original (common-factor) input pattern (on the first four factors) in the computer simulation.

Next, the 100 $\underline{F}$ matrices for each population were transformed to

(i) oblique least squares fits to the appropriate $\underline{P}^*$ matrix, using the approximate reference-structure procedure of Mosier (1939),

(ii) orthogonal least squares fits to the appropriate $\underline{P}^*$ matrix, using the general orthogonal Procrustes method most thoroughly explicated by Schönemann (1966), and

(iii) blind orthogonal simple structure solutions, using the generalized normal orthomax criterion (see Hakstian & Boyd, 1972), with the specification parameter, $\underline{w}$, in the function set to 1 (giving varimax), 2 (equamax), 4, and 16.

For each such transformed solution, $\underline{P}_i$, 12 x 4, an error matrix was established as

$$\underline{E}_i = \underline{P}^* - \underline{P}_i,$$

where $\underline{E}_i$, 12 x 4, is the error matrix for Sample $\underline{i}$ drawn from one of the two populations, and $\underline{P}_i$ is the transformed pattern [according to (i) - (iii) above] for Sample $\underline{i}$. An index of goodness-of-fit was obtained in each case by taking the sum of squared error elements, given by $\underline{tr}(\underline{E}_i'\underline{E}_i)$, where $\underline{tr}$ stands for trace. Thus, the dependent variable in each of the analyses to follow was the quantity $\underline{tr}(\underline{E}'\underline{E})$.

In summary, then, for each of the 100 samples (themselves consisting of 100 "subjects") from the factorially simple data, the unrotated factor loading matrix was transformed to (1) an oblique best fit to the factorially simple $\underline{P}^*$ matrix, (2) an orthogonal best fit to $\underline{P}^*$, and (3) four "blind" orthogonal simple

structure solutions, generating six E matrices [and, hence, six tr(E'E) values] in the process. For each of the 100 samples from the factorially complex population, the same procedure was followed, using the factorially complex P* matrix.

Comparisons were made separately between (1) best-fitting oblique and best-fitting orthogonal solutions, and (2) the four "blind" orthogonal methods. These two separate comparative schemes, in addition, were aimed at answering inferential questions concerning the (1) means (over the 100 samples) of the tr(E'E) values, and (2) variances. The experimental paradigm, thus, was the repeated-measures analysis of variance design, with the sample (n = 100) as the unit of statistical analysis, transformational technique as the "treatment," and tr(E'E) values as the dependent variable. Considering the sample as the unit of statistical analysis may seem a little unusual until one realizes that the phenomenon of factor transformation (or, for that matter, any "treatment" involving manipulation of matrices in which the elements correspond not to observations on individuals, but derivations of these for whole samples) cannot be conceptualized in terms of independent effects on single experimental subjects. Thus, the data sample in the present study consisted of 100 matrices (each based on an n of 100 "subjects") drawn from the simple, and 100 from the complex populations. Since sampling was done with replacement, we treat the problem as one of sampling from an indefinitely large population.

Mean differences in tr(E'E) values for the best-fitting oblique vs orthogonal solutions were assessed--for both the simple and complex data--using the standard t-test for correlated data. Differences in the variances of the tr(E'E) values were assessed for significance using the t-test for correlated variances. Next, mean tr(E'E) values of the four orthomax solutions were compared, for both the simple and complex data, first by means of a simple repeated-measures ANOVA, and lastly by multiple comparisons, using the Tukey procedure. Because no analogous procedure exists for assessing equivalence of $J > 2$ non-independent variances,

RESULTS AND DISCUSSION

For the factorially simple data, the repeated-measures ANOVA on the mean $\underline{tr}(\underline{E}'\underline{E})$ values from the four orthomax solutions yielded an $\underline{F}$-ratio (13.65) that was highly significant using both the apparent test ($\underline{df}$ = 3, 297; $\underline{p} < .00001$) and the conservative (see Winer, 1971) test ($\underline{df}$ = 1, 99; $\underline{p} < .001$). For the complex data, the analogous procedure yielded an $\underline{F}$-ratio (8.26) that was again highly significant (apparent test: $\underline{df}$ = 3, 297; $\underline{p} < .0001$; conservative test: $\underline{df}$ = 1, 99; $\underline{p} < .005$). The results of the subsequent pairwise Tukey comparisons, as well as the correlated $\underline{t}$-tests on the best-fitting oblique $\underline{vs}$ orthogonal, and all the pairwise contrasts on variances are given in Table 1.

Insert Table 1 about here

The results show first of all, that a significantly better fit to the oblique population factor pattern matrix was realized by the oblique, as opposed to orthogonal, least squares transformations. This is hardly surprising and of little interest. The greater variability among the $\underline{tr}(\underline{E}'\underline{E})$ values, however, of the best-fitting oblique than orthogonal solutions--noted with the complex data--$\underline{is}$ surprising and seems to cast some doubt on the belief of greater stability in oblique factor solutions.

Results concerning both mean goodness-of-fit and sampling variability are of interest in connection with the orthomax solutions. To determine whether any differences existed among orthomax solutions for the $\underline{population}$ data, $\underline{i.e.}$, whether in the two populations, one or more of the orthomax criteria just happened to yield better simple structures than the others, the authors rotated the unrotated factor loading matrices for the two $\underline{entire}$ $\underline{populations}$ to orthomax solutions, based on the $\underline{w}$ values used with the sample data: 1, 2, 4, and 16. The respective $\underline{tr}(\underline{E}'\underline{E})$ values (based, as before, on closeness to the appropriate $\underline{P}^*$ matrices) were (1) for the simple data: 1.506, 1.504, 1.502, 1.502, and (2) for the complex: 1.499, 1.460, 1.470, and 1.483. When the magnitude of the differences between

population $tr(E'E)$ values is compared with the magnitude of the mean differences for the sample values--as given in Table 1--it should be clear that the minute population differences can have had no effect whatever on mean differences for the factorially simple data, and almost as certainly had no effect on mean differences for the complex data.

The multiple comparisons among mean $tr(E'E)$ values of the orthomax solutions suggest that, for both factorially simple and complex data, values of the w parameter of around q, the number of factors, or $q/2$ (giving equamax) yield solutions significantly closer to the population parameter pattern matrix, than does either the value 1, which defines the widely-used varimax criterion, or very large w values (for example, $4q$, or 16 in the present study). As has been noted above, this fact cannot be explained on the basis of population characteristics. With regard to sampling variability--assessed by the variances of the $tr(E'E)$ values--again the $w = 2$ and 4 solutions tended to be significantly less variable than both the varimax and $w = 16$ solutions, for the simple data, and less variable than the varimax solutions, for the complex. Thus, overall, the $w = 2$ and 4 solutions tended to be closest to the parameter solutions and least variable of the four orthomax solutions compared.

Three objectives attended the present study. One was to draw attention to the notion of invariance over samples, in connection with factor transformation methods, and to provide some empirical evidence bearing on this notion. The second was to explore the relative fluctuation in solutions derived under the oblique, as opposed to orthogonal, factor models. It has been seen that there is some reason to question any assumption of lesser variability in connection with the oblique model. A final objective was to compare various specializations of the general orthomax criterion, using the standard of sampling invariance in this regard. The results of the present study appear to suggest that values of w in the orthomax criterion in the range of $q/2$ to q yield orthogonal solutions showing relatively high invariance. This fact

may be related to the long-known fact (discussed in Hakstian & Boyd, 1972) that varimax (which is used in the vast majority of all factor studies done) often fails to disperse sufficiently the first general factor, so that if a population simple structure configuration exhibits somewhat level variance contributions among the major factors, varimax may be unable to duplicate this in a set of sample data. As noted earlier by Hakstian and Boyd (1972), for quite different reasons, Saunders's equamax (w = 2 in the present study) may well be a superior "blind" orthogonal technique. Further research is obviously needed in extending the notion of sampling invariance in connection with transformation methods. Studies employing the common-factor, rather than component, model throughout, and "blind" oblique transformation methods would seem most relevant. Sample size, ratio of number of variables to factors, and magnitude of the communalities also may be independent variables of interest.

REFERENCES

Carroll, J. B. An analytical solution for approximating simple structure. Psychometrika, 1953, 18, 23-38.

Hakstian, A. R. A comparative evaluation of several prominent methods of oblique factor transformation. Psychometrika, 1971, 36, 175-193.

Hakstian, A. R., & Boyd, W. M. An empirical investigation of some special cases of the general "orthomax" criterion for orthogonal factor transformation. Educational and Psychological Measurement, 1972, 32, 3-22.

Mosier, C. I. Determining a simple structure when loadings for certain tests are known. Psychometrika, 1939, 4, 149-162.

Schönemann, P. H. A generalized solution of the orthogonal procrustes problem. Psychometrika, 1966, 31, 1-10.

Winer, B. J. Statistical principles in experimental design. (2nd ed.) New York: McGraw-Hill, 1971.

Table 1

Differences[a] between Means and Variances of $tr(\underline{E}'\underline{E})$ Values for the Pairwise Comparisons Made for the Factorially Simple and Complex Data (Correlations Are Given in Parentheses)

Comparison	Means		Variances	
	Simple	Complex	Simple	Complex
Best-Fitting Orthogonal vs Oblique	1.199** (.81)[b]	.906** (.67)	-.016	-.130**
$\underline{w}$ = 1 vs $\underline{w}$ = 2	.105** (.87)	.092** (.97)	.091**	.039**
$\underline{w}$ = 1 vs $\underline{w}$ = 4	.111** (.87)	.090** (.94)	.091**	.034*
$\underline{w}$ = 1 vs $\underline{w}$ = 16	.078** (.87)	.019 (.77)	.056*	.015
$\underline{w}$ = 2 vs $\underline{w}$ = 4	.006 (.99)	-.002 (.98)	.000	-.005
$\underline{w}$ = 2 vs $\underline{w}$ = 16	-.027 (.94)	-.073* (.80)	-.035**	-.024
$\underline{w}$ = 4 vs $\underline{w}$ = 16	-.033 (.95)	-.071* (.82)	-.035**	-.019

[a]A positive difference between means or variances indicates that the mean or variance of the first-noted solution under "Comparison" is larger; a negative difference indicates that the second is larger.

[b]These are correlations among the $tr(\underline{E}'\underline{E})$ values obtained under the two methods given in the comparison.

*Significant at the .05 level; **significant at the .01 level.